

Expert Report by
Emeritus Professor Malcolm Coulthard
and
Professor Tim Grant
Centre for Forensic Linguistics, Aston University, England

3 March 2017

Table of Contents

1.	Introduction	3
2.	Executive Summary	3
3.	Qualifications	5
4.	Materials reviewed	6
5.	Problems with Dr Chaski’s approach	6
	a. Dr Chaski’s methodology has not been scientifically endorsed.....	6
	b. Dr Chaski’s methodology is suspect.....	8
	(i) Dr Chaski’s approach is based on a theory that is unsuitable for authorship attribution.....	8
	(ii) Dr Chaski then builds her model using an arbitrary selection of linguistic and non-linguistic features	9
	c. Nor can Dr Chaski’s statistical method be relied on.....	12
	(i) Dr Chaski does not provide any workings to support her statistical analysis...	12
	(ii) Dr Chaski’s statistical analysis is fundamentally flawed	12
	d. Dr Chaski’s method is based on a series of implausible assumptions about the production of the Awards texts	13
6.	The way in which Dr Chaski has carried out her analysis further disqualifies her conclusions	15
	a. Dr Chaski erroneously attributes authorship of the Awards to Mr Valasek by default.	15
	b. Dr Chaski’s data selection is flawed.....	16
	(i) The texts selected by Dr Chaski to determine authorial style are inappropriate	16
	(ii) Dr Chaski manipulates the selected texts.....	20
7.	Dr Chaski’s analysis produces inconsistent results, further confirming that her analysis is at best unreliable	22
8.	Conclusion.....	23

1. Introduction

1.1 This expert report relates to three arbitral awards (“**Awards**”) rendered in arbitrations initiated by the three former controlling shareholders of the now defunct Yukos Oil Company—Hulley Enterprises Limited, Yukos Universal Limited and Veteran Petroleum Limited (“**Claimants**”) against the Russian Federation. The Awards were issued on 18 July 2014 by a three-member arbitral tribunal, composed of Mr Yves Fortier (the president), Judge Stephen Schwebel and Dr Charles Poncet (“**Tribunal**”).

1.2 The Russian Federation has filed an application to have the Awards set aside before the courts in The Hague, The Netherlands, and has been resisting enforcement by the Claimants of the Awards in other jurisdictions. One of the grounds relied on by the Russian Federation is the claim that the three arbitrators delegated their duty to decide the parties’ case to an assistant, Mr Martin Valasek. In support of this claim, the Russian Federation has produced two expert reports prepared by Dr Carole E. Chaski (“**Chaski Reports**”). In these reports, Dr Chaski claims to have demonstrated conclusively that many subsections of the Awards were written not by the Tribunal members, but rather by Mr Valasek.

1.3 We have been instructed by De Brauw Blackstone Westbroek N.V., counsel for the Claimants in the setting-aside proceedings, and Shearman & Sterling LLP, who advise and represent the Claimants in the various recognition and enforcement proceedings, to analyse the Chaski Reports and the methodology used by Dr Chaski. Professor Coulthard was instructed to examine, in particular, Dr Chaski’s linguistic analysis, while Professor Grant was asked to evaluate her statistical analysis.

2. Executive Summary

2.1 Having reviewed the Chaski Reports, it is our opinion that they do not contain a sound, scientifically defensible and reliable assessment of the authorship of the Awards.

Dr Chaski’s proprietary software and methodology are unverifiable (§§5.1-5.5; 5.15-5.24)

2.2 The proprietary software used by Dr Chaski for attributing authorship of subsections of the Awards has not been made available for review and is therefore untested and unverifiable—data goes into her computer, and results come out, and no-one knows what her software does with the data in between.

2.3 Dr Chaski gives no detailed justification of her methodology. Thus, for example, she gives minimal explanation and scant illustration for her choice of the linguistic and non-linguistic features used to build her model. Likewise, she does not provide the statistical workings on the basis of which she reaches her conclusions. Without this information, it is impossible to properly evaluate the reliability of Dr Chaski’s (linguistic and statistical) methodology: one is left in the dark as to just how Dr Chaski has carried out her analysis. For this reason alone, her analysis cannot be relied on and the results in her reports should be disregarded.

Dr Chaski's linguistic methodology is unsuitable for authorship attribution, especially in the context of collaborative drafting processes (§§5.6-5.14; 5.27-5.34)

2.4 Even without access to Dr Chaski's software or a proper explanation of her methodology, it is readily apparent that she bases her approach on a theory that is not intended, or suitable, for authorship attribution.

2.5 Furthermore, Dr Chaski's methodology fails to account for collaborative drafting processes, like those used to create the Awards. Instead, she **assumes** that the Tribunal must have written all the subsections of the Awards in a common style, which would have resulted in a merging of the individual styles of the three Tribunal members. In doing so, she critically ignores the reality that any given subsection of the Awards would have been produced by one author, with the dominant style of that author then modified by the subsequent interventions of other authors. Dr Chaski's methodology is not equipped to handle such a complex collaborative process.

Dr Chaski's statistical methodology flies in the face of accepted practice (§§6.1-6.3)

2.6 Crucially, there is **no** positive demonstration in the Chaski Reports that Mr Valasek wrote any subsection of the Awards. Rather, Dr Chaski's entire analysis rests on an erroneous binary assumption: namely, that the relevant subsections of the Awards were written exclusively either by (i) the members of the Tribunal (individually or collectively), or (ii) Mr Valasek alone. On that basis, Dr Chaski attributes authorship to Mr Valasek seemingly **by default** each time that, according to her statistical model, the probability that a subsection was written by the Tribunal is low. In other words, instead of running a test to positively attribute text to Mr Valasek, she operates on the (baseless) assumption that if she cannot attribute text to the members of the Tribunal (individually or collectively), then it **must** have been written by Mr Valasek alone. The result is that if a subsection of the Awards which in fact was written by the Tribunal is, due to uncertainty or flaws in Dr Chaski's method, not attributed to the Tribunal, that subsection is incorrectly attributed to Mr Valasek. This is simply wrong, and flies in the face of accepted practice.

Not only is Dr Chaski's initial selection of documents to determine the style of the Tribunal members and Mr Valasek flawed, she then manipulates her chosen data (§§6.4-6.19)

2.7 Dr Chaski selects fundamentally inappropriate texts to determine the authorial style of the Tribunal members and Mr Valasek. Dr Chaski uses documents from genres that are quite obviously different (for example, edited transcripts of speeches) from those of the Awards; ignores the likelihood that the chosen documents have been edited or even partly co-authored by others; and uses documents that were written long before the Awards (despite acknowledging that the style of an author can change over time). All of these issues increase **even further** the chance of her wrongly attributing subsections of the Awards.

2.8 Dr Chaski then manipulates her chosen data, to exclude vast swathes of text of known authorship, on the basis that her approach has failed to attribute them correctly—without stopping to consider how an analytical approach that apparently has difficulty classifying text of **known** authorship successfully can possibly be relied on to classify text of **unknown** authorship.

2.9 By manipulating her data in this way, Dr Chaski claims to have achieved a classification of 100% certainty. This is a clear empirical indication that her statistical model is unreliable.

Dr Chaski's results are inconsistent (§§7.1-7.6)

2.10 In light of all the above, it is not surprising that Dr Chaski's attributions to Mr Valasek in her first and second report are significantly different. These inconsistent results are yet further confirmation (if any were needed) that the Chaski Reports cannot be given any credence.

3. Qualifications

3.1 The Centre for Forensic Linguistics at Aston University (“**Centre**”) combines cutting-edge research and investigative practice with teaching and training in the field of forensic linguistics. The Centre's research involves all aspects of forensic linguistics, including the development and refinement of methods for identifying the author of disputed forensic texts. Its academic staff have published widely and lectured about their research in some 35 countries.

3.2 Professor Coulthard was the Founding Director and is now Emeritus Professor of Forensic Linguistics at the Centre. His *curriculum vitae* is attached as Annex CG-1. By way of summary, he was founding editor of *The International Journal of Speech, Language and the Law*, a peer-reviewed journal created in 1994 as *Forensic Linguistics*, and founding president of the International Association of Forensic Linguists. He is the author of 26 authored and edited books and over 100 academic articles and chapters in books, and has supervised numerous PhD dissertations on various aspects of language and the law. He has been engaged to write expert reports in over 230 cases in 10 countries and has given expert evidence in courts in England, Scotland, Northern Ireland, Germany and Hong Kong.

3.3 Professor Grant is the current Director of the Centre, where he is also Professor of Forensic Linguistics, teaching undergraduate and postgraduate courses in Forensic Linguistics and Research Methods. His *curriculum vitae* is attached as Annex CG-2. By way of summary, his specialised field is forensic linguistics and questioned authorship analysis, including statistical methods for authorship analysis. He has taught statistical methods for postgraduate students in psychology and in linguistics and is lead author on a forthcoming textbook on statistical methods for students of linguistics. He has numerous academic qualifications and publications in this area, as well as extensive experience in providing expert investigative assistance, having given expert evidence for both prosecution and defence in over 100 cases

in England, Scotland, Germany, Canada and Australia. He is the incumbent president of the International Association of Forensic Linguists.

4. Materials reviewed

4.1 In preparing this report, we have reviewed and considered the materials provided to us. In addition to the two Chaski Reports, they include the Awards and the parties' pleadings on this issue in the setting-aside proceedings in the Netherlands. Copies of the Chaski Reports are attached as Annex CG-3 and Annex CG-4, respectively. In this report, we use the following abbreviations:

Expert report of Dr Carole E. Chaski, dated 11 September 2015	CH1
Reply report of Dr Carole E. Chaski, dated 13 January 2016	CH2

5. Problems with Dr Chaski's approach

a. Dr Chaski's methodology has not been scientifically endorsed

5.1 Unlike most other computational scholars involved in authorship analysis (such as Drs Shlomo Argamon, Patrick Juola, Efstathios Stamatatos, to name but three), Dr Chaski does not make her analytic system publicly available for others to use and test¹. Nor has she made this system available to readers of the Chaski Reports. Rather, Dr Chaski's analysis relies on proprietary software devised by Dr Chaski herself. This software, called ALIAS-SynAID, is sold as part of a commercial enterprise.

5.2 Further, Dr Chaski has not published widely or recently in scientific journals in support of her method, meaning that it has not been subjected to the kind of peer scrutiny that is usually considered necessary for recognition as an expert in the field.

5.3 Dr Chaski cites her own publications dated 1997 (CHA-2), 2001 (CHA-3), 2005 (CHA-4), 2005 (CHA-8), 2007 (CHA-5), 2010 (CHA-6), 2012 (CHA-7), 2013 (CHA-9) and 2014 (CHA-10) as providing supporting research to validate her software (CH1-§12). Of these articles, only one, Chaski (2001), is a validation study that appears in an internationally recognised, peer-reviewed, linguistics journal. The 2001 article was heavily criticised by one of us (Grant) in Grant and Baker (2001), as well as by McMenamin (2001). Both Grant and Baker (2001) and McMenamin (2001) concluded that Chaski (2001) is a poorly designed study taking no account of previous literature or approaches and that, as her research contains serious statistical failures, the results of this early research cannot be relied on to support her method or software.

¹ See, as one example of publicly available authorship software, Juola's Java Graphical Authorship Attribution Program, which anyone can download free and evaluate on their own data. See <https://github.com/evllabs/JGAAP>.

5.4 Since the 2001 publication, there have been no published, peer-reviewed follow-up studies by Dr Chaski to correct these flaws:

- Chaski (2005) (CHA-4) provides a series of case studies and further describes a method which was developed from her 2001 article. However, this article is not a validation study of her linguistic methods, and there is insufficient statistical detail to determine whether the method described is any more reliable than the flawed 2001 method. Nor does it set out to address the flaws in Chaski (2001) as identified in Grant and Baker (2001) and McMenamin (2001);
- Chaski (2005) (CHA-8) is a chapter in an edited volume and thus is not blind peer-reviewed. Neither is it a validation study. Rather, it contains a collection of Dr Chaski's own opinions and critiques of others' methods in authorship analysis. She naturally advocates her own methods, relying self-referentially on her 2001 article, a 2004 conference presentation, and 2005 (CHA-4) to suggest her method is validated²;
- Chaski (2007) (CHA-5) is another book chapter, part of which makes some attempt at validation testing, but Dr Chaski does not provide the detail needed for a validation study with regard to either the output from the ALIAS-SynAID software or the statistical methods used;
- Chaski (2010) (CHA-6), another book chapter, is not a validation study;
- Chaski (2012) (CHA-7), a handbook chapter, asserts that ALIAS-SynAID has been validated, but self-referentially cites only Dr Chaski's own 2001, 2005 and 2007 work, all of which has been critiqued above;
- Chaski (2013) (CHA-9) was not peer reviewed. It was published in a student-run legal journal and is primarily a theoretical paper mainly devoted to Dr Chaski's own writings and casework, with a small empirical section, which does not provide a basis for evaluating her method; and
- Chaski (2014) (CHA-10), another law-school journal article, is descriptive and does not provide a validation study.

5.5 As the results of Dr Chaski's early research cannot be relied on and as later publications do not address these criticisms, there is no rigorous scientific support for the use of Dr Chaski's methodology or software. ALIAS-SynAID is not, as claimed, a proven method supported by empirical research. Neither have any independent blind studies been carried out, and Dr Chaski's own studies are not of sufficient quality to support any reliable use of her methodology.

² Dr Chaski also cites two other papers advocating computational methods for authorship attribution in Modern Greek, which have nothing to do with her own methods.

b. Dr Chaski's methodology is suspect

(i) Dr Chaski's approach is based on a theory that is unsuitable for authorship attribution

5.6 Even on the basis of the limited information we have about Dr Chaski's analytic system, we can confidently conclude that it is based on a theory that is unsuitable for authorship attribution. The Chaski model is claimed to be based on a form of syntactic analysis first introduced in 1957 by Noam Chomsky, which "derives from phrase structure grammar, [which is] a way of precisely, and mathematically, stating the rules for combining and ordering words" (CH1-§19).

5.7 A major problem in applying phrase structure grammar to authorship attribution is that it was originally designed to describe the brain's innate knowledge of how phrases are constructed, not to analyse actual texts. Dr Chaski makes a serious mistake in using such a form of analysis for the purpose of authorship attribution and does not even take into account subsequent major developments in cognitive linguistics.

5.8 The reality is that although phrase structure grammar can describe accurately the structure of phrases **out of context**, it is not designed to take account of the way individual phrases are shaped to fit into texts. Phrase structure grammars describe the total set of grammatical options that are available at any given point in a sentence when one is choosing the next word or phrase. Linguists have called this set of options the 'grammatical paradigm'. Chomsky is famous for showing that the same grammar can produce grammatical but meaningless phrases, of which the most memorable is "colourless green ideas sleep furiously". Thus, while such a grammar can successfully describe all the possible structural combinations available, it cannot, and indeed does not set out to, describe how structural choices are limited in and by a given context.

5.9 However, Dr Chaski, unlike Chomsky, is setting out to describe authorial style and so needs to be able to show how different authors make **different choices** when faced with the **same set** of options to choose from. Indeed most linguistic analysis assumes that style is a function of authors making differential choices. For instance, they may choose to express uncertainty with a verb by the choice of 'may' or with an adverb by the choice of 'possibly'.

5.10 Dr Chaski, however, appears to assume that when an author is creating a new phrase at any point in a text, **all** possible grammatical options are available and that the analyst can regard any grammatical choice as an intentional and significant choice. She seems to work on the assumption that whole texts are made up of self-contained sentences, and that sequencing does not affect grammatical choices. This is, of course, incorrect: very frequently, because of **contextual constraints**, the author is unable to choose from all of the grammatical options and is instead limited to a much narrower set of options.

5.11 To take a few examples from Dr Chaski's report: she illustrates the choice between 'unmarked' and 'marked' adjective phrases³ with the following example: "a beautiful truck" is said to be unmarked and "the truck is beautiful" to be marked (CH1-§22). However, let us imagine a possible linguistic context "John walked down the road and he saw ...". At this point the choice of the phrase "the truck is beautiful" is not available; the author can only choose "a beautiful truck". However, if the text had instead continued "John said to his friend 'that car is awful but...'", then the only option is "the truck is beautiful". The author cannot produce "that car is awful but 'a beautiful truck'". Similarly, choosing one verb form rather than another, for example 'bought' rather than 'has bought' (CH1-§22), is frequently not a question of a choice between two or more alternatives and therefore conditioned by a stylistic decision; rather, the use of the one rather than the other of these forms is constrained by context.

5.12 Other internationally recognised computational linguists have therefore moved on to base their analyses on a different form of grammar, called 'Systemic Functional Linguistics' ("SFL"), which sees language as a system of ever-changing options which are constrained by the linguistic and non-linguistic context in which the language is produced. SFL was specifically designed to analyse complete texts, and not just isolated phrases. Among other advantages, it is therefore better suited to answering questions about authorship attribution (see, for example, Argamon and Koppel (2013); Nini and Grant (2013)).

5.13 Other demonstrably successful authorship markers not included by Dr Chaski are the identification of idiosyncratic lexical choices (see Coulthard (2013)); the repetition of three- and four-word sequences known as N-grams (see Johnson and Wright (2014)); and average sentence length (see Winter and Woolls (1993), summarised in Coulthard (2000))⁴. Indeed average sentence length would seem to be particularly relevant in this case, given that Dr Chaski's basic unit of analysis is the sentence (see §§6.7-6.8 below for a discussion of the significance of sentence length).

5.14 On the basis of the unsuitability of the theory on which Dr Chaski's method is grounded, we consider her approach and her results to be invalid.

(ii) Dr Chaski then builds her model using an arbitrary selection of linguistic and non-linguistic features

5.15 In addition to using an inappropriate linguistic theory, Dr Chaski has been selective in her application of that theory to authorship analysis work. Adherence to the linguistic theory of phrase

³ 'Markedness' is traditionally used in linguistics and the social sciences to refer to items that stand out as unusual or rare in comparison to a more common or regular form. It is used to refer to cases where there is a choice; so, for instance, whenever one produces any clause there is always a choice between a positive and negative option: "I am happy" versus "I am not unhappy". The unmarked, positive choice occurs in English texts roughly nine times more frequently than the marked, negative choice.

⁴ The omission of these last two features is particularly surprising as Chaski (2005) asserts that her software is capable of calculating them.

structure grammar suggests a focus on syntactic features. Dr Chaski claims to have chosen to build her analysis on just 22 linguistic features (CH1-§22, [Table 1](#)), representing only a small fraction of those that a detailed phrase structure analysis would actually indicate as potential markers⁵. Having made the choice to work with structural grammar, one would expect Dr Chaski to have included a much larger number of phrase structure features, such as the choice between coordination and subordination⁶, active and passive structures⁷ and uncertainty⁸. These are but three examples of features likely to be important style differentiators in texts produced by sophisticated writers⁹—in particular by lawyers and judges.

5.16 Further, Dr Chaski's statement that her method focuses on 22 linguistic features is incorrect. When one looks at the list of features provided in Table 1 (CH1-§22), the first three are punctuation choices, while the last two are apparently not independent at all, but summations of some of the earlier features. Even more worrying: the 14 grammatical features numbered 4-17 appear actually to be seven reciprocally defined pairs of marked and unmarked choices—for example, any noun phrase must be, and must therefore be classified as, either unmarked or marked, so these 14 features are in reality only 7. In other words, the phrase structure part of the analysis looks at a mere seven independent choices¹⁰.

5.17 To confuse matters further, Dr Chaski has chosen to add to her list of grammatical features a non-grammatical measure, that of word length, which is calculated in terms of the average number of letters per word. Word length is not a grammatical or even a linguistic feature (as there is no linguistic explanation for why it might be distinctive). In addition, Dr Chaski provides no indication of the weight that she gives to average word length in her overall analysis, let alone any examples of the four authors

⁵ In fact only 17 features used by Dr Chaski are phrase structure features; the first three features in Dr Chaski's list are punctuation choices and the final two in the list are summations of **some** of the other 17, although these are not specified or justified.

⁶ Time sequence can be indicated by coordination or subordination. For example: "He switched on the kettle and washed the cups", OR, "**After** he had switched on the kettle he washed the cups".

⁷ If the author does not know or does not want to indicate the doer of an action, the passive verb choice allows this. For example: "He switched on the kettle and washed the cups", OR, "After the kettle **had been switched on** the cups **were washed**".

⁸ If the author wants to indicate degrees of uncertainty this can be achieved through the choice of modal verbs or modal adverbs. For example: "He switched on the kettle and washed the cups", OR, "The cups **must have** been washed **possibly** before the kettle was switched on".

⁹ As seen in footnotes 6 to 8 above, whenever writing, an author is faced with different choices for encoding what is essentially the same message, and authors can often be distinguished by which choices they prefer. Let us now take a real example and focus on the final 10 words of the previous sentence in this footnote: "and authors can be distinguished by which choices they prefer". In linking the clause to what had gone before, we chose a simple coordinator, 'and', where we could have indicated a causal relation by using 'so'; we chose a passive construction, 'authors can be distinguished', which meant we did not need to say who could do the distinguishing; and we modalised the whole clause with 'can often', to indicate that however good a method is, there will be times when it is unsuccessful.

¹⁰ If Dr Chaski has actually treated the members of these pairs as if they were independent choices, this has devastating consequences for the use of the DFA statistical package, which requires all features to be independent of each other (see [Annex CG-5](#)).

differing in terms of word length and so being distinguishable in both the known and the disputed texts. Finally, as she has included word length in her analytical system, there is no apparent reason for the exclusion of other well-tested non-grammatical authorship markers such as the length of sentences and the occurrence of repeated sequences of letters or words (see above, §5.13).

5.18 The importance of having a principled basis for feature selection cannot be over-estimated. Dr Chaski cannot know whether the inclusion of additional features or of different features would have changed the outcome of her analysis. Her feature selection was driven mostly by an inappropriate theoretical assumption that phrase structure grammar can provide a sound basis for the selection of features in an authorship analysis. Having set out and attempted to defend phrase structure grammar as her position, Dr Chaski then in an *ad hoc* manner omits to use potential features suggested by such a position and yet includes other features, like punctuation and word length, which would fall outside that position.

5.19 The small ragbag of linguistic features that Dr Chaski has chosen to rely on for her analysis are furthermore given no justification and minimal illustration. For instance, one has no idea what even the basic category of ‘pronoun’ means for her—does she count the frequency of all pronouns, or just of personal pronouns in subject and object position; and is ‘his’, for example, always categorised as a pronoun, or is it classed as some kind of adjectival or determiner in phrases like ‘his red car’ and perhaps as a nominal in ‘this is his’? Similarly, we have no idea if Dr Chaski programmed her software to calculate the percentage of times a pronoun is chosen instead of a noun, or if she simply counts raw frequencies, nor whether pronouns can (like other of her grammatical features) be classed as either marked or unmarked and if so what that would mean¹¹. None of these important decisions about definition and implementation are discussed, explained or exemplified, but these choices will nevertheless have a major impact on the outcome of Dr Chaski’s analysis.

5.20 Dr Chaski does not examine, let alone justify, in her writings why her chosen linguistic features should vary between authors. Nor is there any broader research literature examining features from phrase structure grammar for the purposes of attributing authorship. Without a full description and discussion of the distribution of each of these features and a demonstration of why they might be expected to vary systematically, there is a very real danger of statistical misclassification. For example, without proper discussion, explanation and understanding of how each feature fluctuates as a variable according to the nature and functions of a text, the danger is that any statistical classification will in fact be partly or even wholly dependent on the type of text, rather than the authorship of the text¹².

¹¹ On the meaning of “markedness” in linguistics, see footnote 3 above.

¹² This is why we evaluate below (in **Section 6.b(i)**) the validity of Dr Chaski’s assumption that an analysis of the style of a small set of academic articles can be used in order to characterise the style of individual subsections of the Awards.

5.21 Nor does Dr Chaski provide any illustrative examples taken from any of the texts involved in the present case (neither from the candidate authors' known academic articles, nor from the disputed Awards), which could show how her chosen linguistic features distinguish between the candidate authors. Instead, she uses what appear to be invented examples to illustrate the features.

5.22 To conclude, Dr Chaski's method is not based on rigorous linguistic analysis; she uses a very narrow, arbitrary assortment of features, the individual weights of which are not disclosed and the attributive success of which has not been demonstrated in publications exposed to peer review.

c. Nor can Dr Chaski's statistical method be relied on

(i) Dr Chaski does not provide any workings to support her statistical analysis

5.23 It would be normal practice in writing a linguistics expert report to provide workings that are sufficiently detailed to be evaluated by a second expert. In authorship analysis work this would typically include: characterisation of methods; if, as here, computational methods were used, code for feature extraction; tables of the feature counts broken down by text; details of the statistical software used; the input and output files used in the statistical analysis; and finally a summary of the more significant and also minor statistical design decisions taken. None of these expected workings have been provided.

5.24 The purpose of providing such workings is to allow for the possibility of evaluation by replication of results. A responding expert should be able to check whether the feature counts are accurate, and indeed whether the computational method counts all and only the intended features. We have noted above that the ALIAS-SynAID software is not available for independent evaluation and/or replication studies and that this is a serious deficiency in Dr Chaski's method. Further, there is an equivalent lack of transparency in Dr Chaski's application of statistical methods. Her calculations cannot be evaluated or re-run and the myriad of minor choices that are made in any statistical analysis are not documented. In effect, Dr Chaski seems to expect that it will be blindly trusted that she is a statistical expert, that she has carried out her analyses competently and that she has made defensible choices in statistical design and application. Nothing in the Chaski Reports provides grounds for such blind deference.

(ii) Dr Chaski's statistical analysis is fundamentally flawed

5.25 In spite of the lack of detailed statistical workings, it is still possible to comment on the statistical method that Dr Chaski has chosen. Dr Chaski uses the numbers produced by the ALIAS-SynAID software to carry out a form of multivariate statistical classification known as Discriminant Function Analysis ("DFA"). This means that Dr Chaski has taken a number of variables (in this case, the linguistic and non-linguistic features she considers to be reliable authorship markers) and has combined them into mathematical formulae, which she then uses to discriminate between groups of data points (which in this case are the subsections of the Awards).

5.26 DFA is designed to solve classification problems, and can be used in authorship analysis to classify texts by author. Indeed, when used properly, DFA is a powerful method. However, in order to produce reliable results it requires a series of statistical assumptions to be met. It is *a priori* very unlikely for this to be the case for linguistic data, as explained in more detail in Annex CG-5 to this report. Given that she has not provided any of her statistical workings, it is not clear how, if at all, Dr Chaski checked that the variables she chose can be used in a DFA. Her statistical analysis is therefore fundamentally flawed.

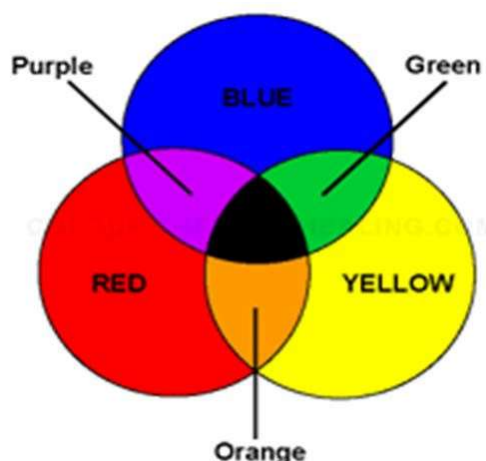
d. Dr Chaski's method is based on a series of implausible assumptions about the production of the Awards texts

5.27 In addition to the problems with her linguistic and statistical analyses described above, Dr Chaski also ignores the crucially important fact that creating a document like the Awards is an essentially collaborative task. Specifically, in applying her method, Dr Chaski assumes:

- that each subsection of the Awards text was exclusively produced, drafted and edited by (i) Mr Valasek, (ii) all three Tribunal members collectively (CH1), or (iii) one of the Tribunal members individually (CH2);
- that no other persons played any role in producing, drafting, or editing the Awards texts; and
- that where a subsection of the Awards texts is taken to have been produced by all three Tribunal members collectively, then the proportion of the individual contributions of each Tribunal member in production, drafting and editing remains constant across all sections of the Awards.

5.28 In reality, however, the text of every subsection would have been produced in the first instance by one author and would therefore be in the style of that author. The style would then have been modified to an unknown degree by the subsequent interventions of other authors. The degree to which it was modified would of course depend on the extent of their additions, deletions, re-phrasings, etc.

5.29 Let us use a simple colour analogy to explain this. Imagine there are three authors, Mr Blue, Mr Yellow and Mr Red. Virtually all of the first draft of the Awards would be in one of the three primary colours. As the two other authors subsequently made alterations to Mr Blue's text, some of it would become greenish, some purplish and, when both of these authors edited the same piece of text equally and severely, it would move towards and sometimes actually become a homogenised black, as in **Figure 1** below, indicating equal tri-authorship.

Figure 1: Colours used as a metaphor for the joint authorship of the Awards

5.30 This illustration is, of course, still a massive simplification for situations like the present case, where, in reality, more than three people could have made edits to the draft of all of the subsections of the Awards, thus altering their style even further. For instance, the various drafts of the Awards may have been proofed by staff from the Permanent Court of Arbitration, who would necessarily have added their own smattering of punctuation preferences and grammatical and lexical idiosyncrasies. In addition, there is the possibility (also overlooked by Dr Chaski) that a staff member (or staff members) of the Permanent Court of Arbitration drafted those subsections of the Awards which simply summarise the parties' positions, and that those subsections would in turn have been modelled on the parties' own submissions, introducing yet further foreign linguistic markers. This would further impact the collaborative style of the Awards texts. To use the same colour analogy as above, four, five or possibly even more colours were potentially involved, and mixing them would result in hundreds of different nuanced hues.

5.31 Dr Chaski's methodology is not designed to take into account this collaborative process. In fact, she makes no attempt at all to deal with the issue. Instead, she oversimplifies her task, first into a two-author and then into a four-author question.

5.32 Dr Chaski initially assumed that "given that the Final Awards were ostensibly co-authored by the Tribunal members, [she] would expect to find a style that merges elements from all three members' individual styles, which can be approximated by considering the writings of the Tribunal members together" (CH1-§39). In other words, she assumed that all the text authored by the Tribunal members would be in a uniform black, resulting from a mix of the three arbitrators' styles that remains constant throughout the Awards. As we can see from **Figure 1** above, there is absolutely no reason at all to assume this. Instead of expecting one "style that merges elements from all three members", one would expect a myriad of styles each merging the "elements from all three members" in differing proportions. Thus, the homogenised system Dr Chaski used in her first report will, in reality, only be fully appropriate to

categorise a tiny proportion of the actual Awards¹³. Dr Chaski's simplification of the process is a very unusual step to take in authorship attribution, as it effectively means comparing an 'average', collective style of the Tribunal with the style of one individual, Mr Valasek.

5.33 In her second report, following criticism of her approach by the Claimants, Dr Chaski undertook a red, blue and yellow single-author analysis and this gave significantly different results, with much more text now being assigned to the Tribunal members (see further below, §§7.2-7.3), but she apparently did not go on to test again the unattributed residue with the black system, that is assuming some composite authorship, as she had done in her first report. There is thus a gap between her two reports. There is no indication that she might have even considered that she would need a system to cope with the purple, green and orange styles, and all the other shades of sub-variation in between. Only then could Dr Chaski have reasonably claimed that there was text that she could not assign to the Tribunal members. Such a mixed-author analysis has long been possible. Winter and Woolls, as long ago as 1993, devised a simple analytic system which not only successfully distinguished text written by Author A from text written by Author B, but also distinguished both from jointly authored text.

5.34 Thus, Dr Chaski's assumptions about the production of the awards are overly simplistic and hence implausible. This compromises her method of authorship attribution and renders the results worthless.

6. The way in which Dr Chaski has carried out her analysis further disqualifies her conclusions

a. Dr Chaski erroneously attributes authorship of the Awards to Mr Valasek by default

6.1 Most forensic linguists in the field of authorship attribution are careful to work with a category of 'Unable to Assign Authorship' or 'Unknown Author', so that if the evidence is not strong enough to assign to any of the candidate authors, they explicitly state their inability to do so.

6.2 However, by choosing DFA, Dr Chaski explicitly excludes this as a possible categorisation option. This is because DFA as an analytic tool requires the assumption of a closed set¹⁴; in other words, when used for authorship attribution, DFA determines on the basis of a pre-selected list of authors which is the

¹³ In addition, even the 'average style' inferred by Dr Chaski was characterised on the basis of an analysis of selected portions of known articles and samples of very different sizes for which no statistical adjustment was reported. This reduces the likelihood that any piece of text produced collaboratively by the Tribunal members will be correctly attributed, and increases the likelihood that it will therefore be wrongly attributed to Mr Valasek. For instance, Judge Schwebel's text sample was already the smallest: it initially consisted of some 8,400 words—compared with Dr Poncet's much greater total of 13,800. However, Dr Chaski then classified 36% of Judge Schwebel's writing as 'outliers', and so his total was reduced to some 5,400 words—only 40% of the size of Dr Poncet's (on the issue of 'outliers', see below, §§6.16-6.19). It appears that Dr Chaski does not even attempt to explain how this will not result in Dr Poncet's style dominating the composite Tribunal style, probably because she cannot.

¹⁴ The issue of the closed set assumption is further discussed in Annex CG-5. This issue was explored in Grant (2007), which suggests methods by which these forced (mis)classifications in DFA could be identified and evaluated.

most likely to have written the text. So, even if the style of a particular text is very different from the style of all of the pre-selected authors, authorship would nevertheless be attributed and attributed to the author whose style is the most similar (or the least dissimilar). Thus, in her first report Dr Chaski forces a binary decision: authorship is attributed either to the Tribunal or to Mr Valasek. In the second report, she tests a closed set of just four possible individual authors (Mr Valasek and the three members of the Tribunal individually). In both cases where Dr Chaski's method cannot attribute authorship positively to the Tribunal members, it attributes authorship **by default** to Mr Valasek. Such a simplistic assumption is incorrect, particularly, as explained above (**Section 5.d**), in the cases of collaborative texts, such as the Awards. Indeed, if a subsection of the Awards which in fact was written by the Tribunal is, due to uncertainty or flaws in Dr Chaski's method, not attributed to the Tribunal, that subsection is incorrectly attributed to Mr Valasek. This is simply wrong, and flies in the face of accepted practice.

6.3 After failing to classify some of the text as having been written by the Tribunal members, a possible strategy for Dr Chaski might have been to conduct a second analysis to discover which of those subsections that could not be positively assigned to the arbitrators could be positively assigned to Mr Valasek. This might have mitigated this flawed methodology as at this point, any subsections that could not be confidently assigned to one of the four candidate authors would have been put into a third category. This, of course, would not demonstrate that any or all of these subsections were not produced by one of the two (or four, or more) candidate authors, but simply that the method was unable to decide with confidence which of them was (were) the author(s).

b. Dr Chaski's data selection is flawed

(i) The texts selected by Dr Chaski to determine authorial style are inappropriate

6.4 Before she can analyse the Awards, Dr Chaski must first set up her analytic system by using documents of known authorship. She writes: "For known documents, I used a subset of the numerous articles in the public domain attributed to the Candidate Authors" (CH1-§31). She later defends this selection in her second report: "In any event, it is apparent to me, and I believe it would be apparent to any expert linguist, that the Candidate Authors' academic articles are very similar to the Final Awards in terms of genre" (CH2-§9). In fact, it would not "be apparent to any expert linguist" that the two sets of texts would be "very similar ... in terms of genre"—indeed most linguists would assume that they would be significantly different, that is, that they belong to dissimilar text genres. Every piece of communication, be it spoken or written, is designed for a target audience. The Awards are aimed at legal professionals in the field of arbitration, and use long, grammatically complicated sentences¹⁵. They have genre-specific layout and punctuation conventions and highly specialised vocabulary. Journal articles and even more so

¹⁵ Dr Chaski's own figures show that 40-word sentences are common, and 50-word sentences by no means uncommon, in the Awards. See further below, §§6.7-6.8.

speeches, by contrast, are addressed to a more general audience; the sentences are shorter, the grammar less complicated and the vocabulary less formal. Dr Chaski is not comparing like with like¹⁶.

6.5 Dr Chaski also states that, in selecting the articles for comparison purposes, she “avoided documents that appeared to be lightly edited transcripts of speeches” (CH1-§31), because these deviate strongly in terms of style from an arbitral award; yet even so, two of the three publications chosen to represent Mr Fortier’s style are in fact edited versions of speeches. Even if these were based on texts that were originally written to be read aloud, they will still be significantly different stylistically from texts originally written to be read silently¹⁷. It is well accepted that speech is less ‘dense’ than written text—it has a simpler grammatical structure and a lower proportion of lexical words to grammatical words, which is why text that was originally written to be read silently is more difficult to understand when read aloud than is speech composed in real time, and why there is a real skill in writing speeches, that is, in producing text written specifically to be spoken. Indeed, we do not even know if Mr Fortier produced some or all of the text of his original speeches, or if it was written by an assistant, or indeed if there was any text at all. The published text might also have been created by someone else who transcribed and polished an audio-recording of the spoken version and submitted a written text for approval by the ‘author’. This has certainly happened to Coulthard more than once. Thus, the appropriateness of the some of the selected articles as comparison texts is further compromised.

6.6 A striking example of inappropriate data selection is Mr Valasek’s 358-word *Acknowledgements* ‘article’. Short introductions to journal issues are very different in kind from even the academic articles they introduce, and totally unlike subsections of arbitral awards¹⁸. Despite having been selected by Dr Chaski to sit alongside the other articles, the *Acknowledgements* is quite clearly not an academic article. Further, when we examine this text there is a serious problem, specific in fact to most Acknowledgements. In fact, almost one-third of all the words in Mr Valasek’s *Acknowledgements* (some 113 of the 358 words)

¹⁶ In response to an objection raised by the Claimants that academic articles were inappropriate comparison texts, Dr Chaski asserts in her second report that “even if the articles were from a different genre or were written in a different register as compared to the Final Awards, my reported research has shown that my method can work robustly when the known and questioned documents do not share the same genre or register” (CH2-§9). She even suggests that love letters could have been used. However, the only evidence offered to support this particularly contentious claim is research reported in Dr Chaski’s own publications.

¹⁷ For instance, Fortier (2009) begins conversationally: “**I am both honoured and flattered** to participate in this event which serves as the ‘official celebration’ in Germany of the golden anniversary of BITs...”.

¹⁸ In this respect, it is useful to present, at this point, a short example of an Awards paragraph, which Dr Chaski assigned to Mr Valasek, to be compared with the text of Mr Valasek’s *Acknowledgements* reproduced in full below as footnote 19:

“1258. In its Counter-Memorial, Respondent renews its objection that the Tribunal lacks jurisdiction pursuant to Article 26(3)(b)(i) of the ECT. In its view, developments subsequent to the Interim Awards show that Claimants, as Yukos shareholders, are seeking before the ECtHR damages for the same alleged loss arising from Yukos’ demise. Accordingly, Respondent submits that the ECT arbitrations expose it to double recovery.”

occur in borrowed, pre-formed phrases¹⁹, which he simply incorporated into his text, rather than in phrases specifically created by Mr Valasek. It must be remembered that Dr Chaski assumed that Mr Valasek made all the linguistic choices in this text.

6.7 One issue that should have alerted Dr Chaski to potential problems is the massive difference in average sentence length between these academic articles and the Awards texts. As noted above, Dr Chaski, although accepting word length as one of the 23 items she sets out to measure, has not adopted sentence length as a feature, although it is widely acknowledged to be equally successful as a marker of authorship within genres (see Coulthard (2000) for a report of two authorship attribution cases involving its use). Had Dr Chaski chosen to calculate sentence length, she would have been immediately suspicious of her own claim of stylistic compatibility between the academic articles and the Awards texts. In the articles, the average sentence length—which can be derived quite easily from Dr Chaski’s own figures for total words and total number of sentences presented in Tables 2 and 5 in CH1 (see CH1-§§34 and 41)—is 28.67 and 28.75 for the Tribunal, and 28.03 and 27.39 for Mr Valasek.

6.8 However, when we look at the average sentence length for the Awards (CH1-Annex A), we notice immediately that it is very much greater, at 35.06 words, suggesting a very different kind of writing. Even more worrying, the first and last subsections of Section IX of the Awards, both incidentally credited to Mr Valasek by Dr Chaski with complete confidence—100% probability—have an average sentence length of 45 and 65 words, respectively. Further, three sequential subsections of the Awards (Section X, E3a, E3b and E3c) have collectively 29 sentences with an **average** sentence length of 48.5 words. Faced with these figures, Dr Chaski should at least have questioned whether she had chosen the correct type of comparison

¹⁹ All pre-formed items (titles and proper names) have been indicated in bold in the full text below:

“On behalf of the **McGill Law Journal** and all its members, I would like to thank the many people who made possible this special issue commemorating **Francis Reginald Scott**. This issue sprouted from ‘**The Blasted Pine**’, a conference held one year ago at **McGill University’s Faculty of Law** to celebrate the inauguration of the **F.R. Scott Chair in Constitutional and Public Law**. **Roderick Macdonald**, first **F.R. Scott Professor**, has contributed his inaugural lecture and a paper he presented at the conference. The issue’s other principal articles, by **Harry Arthurs**, **Alan Cairns** and **Jennifer Nedelsky**, are also descended from the papers that each presented at ‘**The Blasted Pine**’. For the conference’s great success, and for its contribution to this issue of the **McGill Law Journal**, I would like to thank its organizing committee (**David Stevens**, **Daniel Jutras**, **Roderick Macdonald** and **Jeremy Webber**) and acknowledge the financial support for the conference provided by **The Richard Golick Endowment Fund** and by **McGill Law’s** Classes of ‘63, ‘75 and ‘77. We also owe a debt to **Dean Stephen Toope** for planting the idea for ‘**The Blasted Pine**’. Serendipitously, it was **Dean Toope** who pronounced the most recent **F.R. Scott Lecture** (in May 1996), an annual tribute in memory of the poet-scholar sponsored by the **Friends of McGill Libraries**, and I thank the Dean for contributing his talk on ‘**Cultural Diversity and Human Rights**’ to this issue. I also thank **John Whyte** for his review of **Samuel LaSelva’s The Moral Foundations of Canadian Federalism**, and **Michael Oliver** for his insightful foreword. Delightfully, some of **Frank Scott’s** poems appear in the included texts. The **McGill Law Journal** thanks **McClelland & Stewart**, publishers of **The Collected Poems of F.R. Scott**, for granting us permission to use them. We are also very grateful to **Peter Dale Scott**, **Frank Scott’s** son and also a poet, for providing an excerpt of his work in progress ‘**Fire-Tending in the Land of Medicine Buddha**’. His stanzas are a beautiful and touching personal tribute to this giant of our Faculty. Scott’s bold and resilient spirit, as evidenced in the pages of this issue, lives on and continues to do good work.”

data and tried to explain how such a massive difference is not significant in this particular case. Dr Chaski does neither.

6.9 Aside from the fundamentally mistaken assumption that academic articles and the Awards are of the same genre, the text selection made by Dr Chaski is problematic for other reasons. Let us explore these problems further.

6.10 Firstly, Dr Chaski assumes that single-author academic articles provide direct unadulterated access to an author's individual style. In reality, academic and copy editors can interfere significantly with the style of published articles. See the extract below from the beginning of a submission that one of us (Coulthard) happens to be currently editing for his own journal—all the items highlighted in red are his editorial additions and those in purple his deletions:

In Common Law systems a general principle is that a person accused of a serious crime is found guilty or not guilty by a **randomly selected** team of ordinary citizens, called a jury. The jury gives the verdict – **that is, it decid**~~ing~~**es** what crime has been committed, **and whether the accused is innocent**~~ee~~**or guilty**. The judge decides the penalty. In reality, things are more complicated, but this gives the general situation.

At various stages in a jury trial, (particularly the beginning and the end), the judge must explain **to the jury** the law that applies to the case, and instruct **them** ~~the jury~~ on the way they **should** ~~view~~**-evaluate** evidence and witnesses.

6.11 It can be seen that among other changes made, Coulthard has altered punctuation, deleted a noun phrase and a verb, introduced conjunctions, pronouns and a modal verb and added 16 words, some 14% of the new total. Such editorial interventions would necessarily affect substantially the output of a Chaski analysis of this particular author's style²⁰. She does not consider, even to dismiss, the possibility that even a few of the academic articles chosen to represent the four authors' individual styles might have suffered similar editorial intervention.

6.12 In addition, Dr Chaski states that she “did not, for obvious reasons, analyze articles that a Candidate Author co-authored with someone else” (CH1-§31). This decision was taken in an attempt to ensure that the texts represented only the style of the named author. However, just as Dr Chaski has ignored the possible influence of editors, so she fails to acknowledge the possibility that articles in which the author expresses his ‘thanks’ to a young researcher or colleague are not the sole product of that named author, but may have been edited or even partly (co-)authored by the collaborator. For example, Mr Fortier, in each of the three articles selected by Dr Chaski, thanks a young researcher or colleague.

6.13 Dr Chaski additionally notes that: “To maximize the accuracy of my analysis, I included articles that were written by the Candidate Authors as recently as possible in relation to the Final Awards” (CH1-

²⁰ Similarly, Dr Chaski's analysis of the Awards is predicated on the (erroneous) assumption that the subsections she has analysed are single author productions (on this point, see further above **Section 5.d**).

§31). If there were indeed “numerous articles in the public domain attributed to the Candidate Authors” (CH1-§31), why does she include for Fortier, Schwebel and Valasek himself only one that had been written in the previous five years—Fortier: **2015**, 2009, 1999; Schwebel: **2015**, 2005, 2003, 1998; and Valasek himself: **2012**, 2010, 2007, 1997? Indeed, five of the chosen 14 articles were at least 10 years old, with three dating from the 1990s. A rapid internet search shows that Judge Schwebel is in fact a prolific writer and there are many more recent publications she could have chosen for Mr Valasek to avoid using the very short, 20-year old, *Acknowledgements*. This careless data selection further weakens the validity of Dr Chaski’s results.

6.14 It is evident from the above that the texts Dr Chaski chose to try to establish the distinctive style features of the four candidate authors were totally inappropriate for the purpose: academic articles clearly do not belong to the same genre as the Awards; some of the articles were originally produced in spoken form; the degree of input by other authors including editors was unknown; the vast majority of the chosen articles were not written contemporaneously with the Awards; and one is a very short Acknowledgements text.

(ii) Dr Chaski manipulates the selected texts

6.15 Once Dr Chaski had collected, to her satisfaction, articles of known authorship, she set out to test the success of her set of features in classifying the authors. In order to do so, she subdivided the 14 articles into 63 smaller units which she labelled ‘sections’²¹.

6.16 Dr Chaski states that her model, created by the DFA, initially had an error rate of 17.5%; that is, 11 of the 63 sections were not assigned to the known author, but rather mis-assigned to a different author (CH1-§40). Dr Chaski apparently did not think this lack of success raised serious questions about the viability of the method itself. So, instead of either changing or at least refining the component feature set of the analytic system, so that it would cope better with the data that she had collected, or collecting more appropriate data, Dr Chaski took the unusual step of ‘refining’ the data she had already collected, so that it would fit better with her analytic features and thereby improve the classification rate. She did this by simply removing the 11 misclassified sections whose authorship was, of course, already known. She justified this procedure by labelling the style of the omitted sections as ‘atypical’, or to use her ‘scientific’ term, ‘*outliers*’. She writes: “These 11 sections are exceptions or outliers with respect to the authors’ individual styles and can be omitted from the analysis in this case because there is sufficient, even excessive, data to run the analytical method without including these texts” (CH1-§41). She reports that

²¹ Where articles had already been divided into reasonably sized sections by the author himself she accepted these divisions; but when there were long uninterrupted stretches of text, she herself subdivided the text into sections roughly 1,000 words long (CH1-§§32-33).

once she had done this, “The resulting model produced a cross-validated accuracy rate of 100% for differentiating between the known documents of the Tribunal and Mr. Valasek” (CH1-§42).

6.17 When one examines what Dr Chaski has labelled as outliers, it becomes evident that this is a totally unacceptable procedure. On examining in detail the table in CH1-Annex A, which lists every one of the data sections that Dr Chaski analysed and indicates with asterisks those that were subsequently excluded, it becomes clear that there is a major conceptual problem. Dr Chaski asserts that she had selected single-author articles in order to avoid contamination, but then she removes significant amounts of text from articles whose authors were already known, without attempting to provide any explanation at all of how the authors came to write so much text that was ‘atypical’ of their normal style. These exclusions are particularly flagrant in the cases of Judge Schwebel and Mr Valasek. When we look, for example, at the breakdown of what was removed from Judge Schwebel’s output, we discover that Dr Chaski has classified the first two-thirds of one article and the entire second half of another as ‘atypical’. These ‘atypical’ sections constitute 67% and 48% respectively, of the whole articles.

6.18 If, as Dr Chaski claims, some 3,000 words, removed from two complete single-author articles which jointly totalled only 5,079 words, can be said to be ‘atypical’, when it is already known that they were written by the author, three major questions arise: first, how can an author writing unaided produce the majority of the text of two separate articles in an ‘atypical’ style; second, what chance does a method that cannot even classify known text correctly have of classifying unknown text correctly; and third, how do we know that what was classified as ‘atypical’ style is not in fact the style that Judge Schwebel later used to write his Awards text, text which Dr Chaski’s program would then obviously wrongly classify as having been written by Mr Valasek²²? No answers are provided by Dr Chaski. She ignores these obvious problems with her method, which, irrespective of deficiencies in other areas of her methodology, cumulatively cast serious doubt on the reliability of her results.

6.19 In addition to the linguistic and logical issues there is a real problem from a statistical perspective with Dr Chaski’s removal of what she labels ‘outliers’. First, it is important to be clear that these excluded text sections appear not to be statistical outliers which have been identified mathematically²³. As explained in §6.16 above, Dr Chaski appears to use the term non-technically to mean simply ‘misclassified cases’²⁴. This in itself throws doubt on the method, which demonstrably cannot correctly assign texts to their already-known author. Second, and as noted above, Dr Chaski’s removal of these atypical cases is said to

²² It is important to note also that even Mr Valasek is judged to produce atypical, outlier text. Indeed not only was almost a quarter of his sample also judged atypical, but one complete 1,000-word article was excluded as an outlier.

²³ Statistically outliers would normally be defined as data points at some specific distance from the mean. This is explained further in Annex CG-5.

²⁴ Once again, the lack of detailed statistical workings make it unclear whether Dr Chaski would distinguish between misclassified texts that are and are not mathematical outliers.

produce an analytical system with a cross-validated accuracy of 100% correct classification (CH1-§42). While this claim of 100% accuracy may sound like a good result, it is, in fact, very unlikely that a robust DFA will ever produce any classification of 100% accuracy, and such a result is usually referred to as the product of an ‘over-fitted’ model. Over-fitting essentially means that a generalisation has been made on the basis of insufficient examples (or insufficiently varied examples), so that the statistical model cannot be used to classify new cases. In other words, the 100% accuracy that Dr Chaski claims her analysis achieves is not an indication of the reliability of this analysis, but rather the opposite: it suggests that her analysis is carried out on the basis of insufficient or inadequate data and is therefore unreliable. So, by excluding the misclassified texts Dr Chaski has not strengthened her analysis—she has weakened her classification model by over-fitting.

7. Dr Chaski’s analysis produces inconsistent results, further confirming that her analysis is at best unreliable

7.1 Dr Chaski’s own results provide further strong evidence that her conclusions about the authorship of the Awards are totally unreliable.

7.2 In her first report, Dr Chaski set out to assign authorship of subsections of the Awards either to the Tribunal members collectively or to Mr Valasek. She asserted that Mr Valasek wrote 80 of the 112 subsections, as shown in **Figure 2** below.

Figure 2: Table 6 of CH1, assigning to Mr Valasek 80 out of 112 subsections

	Section IX <i>Preliminary Objections</i>	Section X <i>Liability</i>	Section XII <i>The Quantification of Claimants’ Damages</i>	Total
Valasek	78.57% (22/28)	65.38% (17/26)	70.69% (41/58)	71.43% (80/112)
Tribunal	21.43% (6/28)	34.62% (9/26)	29.31% (17/58)	28.57% (32/112)

7.3 In her second report, she set out to assign subsections individually to the Tribunal members or to Mr Valasek. Had the original attributions in the first report been correct, the subsections assigned to Mr Valasek would have remained constant, while those assigned to the Tribunal collectively would have been re-distributed among the individual Tribunal members. However, that did not happen. Mr Valasek was now said to have written 13 fewer subsections and the Tribunal members collectively 13 more, an increase of some 40%, as is shown in **Figure 3** below.

Figure 3: Table 8 of CH2, assigning to Mr Valasek 67 out of 112 subsections

	Section IX <i>Preliminary Objections</i>	Section X <i>Liability</i>	Section XII <i>The Quantification of Claimants' Damages</i>	Total
Valasek	67.86% (19/28)	53.85% (14/26)	58.62% (34/58)	59.82% (67/112)
Fortier	21.43% (6/28)	15.38% (4/26)	12.07% (7/58)	15.18% (17/112)
Poncet	0.00% (0/28)	11.54% (3/26)	13.79% (8/58)	9.82% (11/112)
Schwebel	10.71% (3/28)	19.23% (5/26)	15.52% (9/58)	15.18% (17/112)

7.4 There was no acknowledgement of this change of attribution, let alone an explanation.

7.5 Further, unlike in her first report (CH1, Annex B), Dr Chaski did not indicate in her second report *which* specific subsections had been (re-)assigned to individual Tribunal members and to Mr Valasek, respectively. That renders impossible any proper analysis of the numerous (re-)assignments that result from Dr Chaski's methodology.

7.6 To conclude, the minimum requirement of any authorship attribution method is that it provides the **same** answer to the **same** question on **different** occasions; in this case to the question "which subsections did Mr Valasek write?" As is evident, Dr Chaski cannot meet even this minimum requirement.

8. Conclusion

8.1 We have demonstrated that there are flaws in every aspect of the Chaski Reports. Her linguistic and statistical models are inappropriate for authorship attribution; she does not take into account the collaborative process of writing the Awards; comparison data are badly chosen; her method for establishing the style of the three arbitrators and the Tribunal's assistant on the basis of the selected texts is fatally flawed; and her fundamental premise for authorship attribution—that text which cannot be attributed to the Tribunal should be attributed to Mr Valasek—is simply a misinterpretation of the statistical output. It is unsurprising that the Chaski Reports produce inconsistent results.

8.2 To conclude, the two Chaski Reports cannot be given any credence.

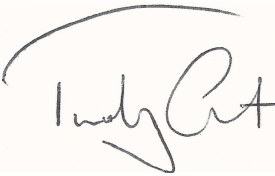
8.3 The opinions that are expressed herein are our own. They are in no way endorsed by Aston University. We declare that we have made all the inquiries we believe are necessary and appropriate and to our knowledge no relevant matters have been omitted from this report; that the facts within our knowledge that have been stated in this report are true; that the opinions expressed in this report are independent and impartial; and that we have complied with the requirements of the Code of Practice of the International Association of Forensic Linguists.

Emeritus Professor Malcolm Coulthard

Signed: 

Date: 3 March 2017

Professor Tim Grant

Signed: 

Date: 3 March 2017

List of Annexes

No.	Description
CG-1	<i>Curriculum vitae</i> of Emeritus Professor Malcolm Coulthard.
CG-2	<i>Curriculum vitae</i> of Professor Tim Grant.
CG-3	Expert report of Dr Carole E. Chaski, dated 11 September 2015.
CG-4	Reply report of Dr Carole E. Chaski, dated 13 January 2016.
CG-5	Grant, T., Note on the discriminant function analysis and its application in authorship attribution.
CG-6	Argamon, S. and Koppel, M. (2013) “A Systemic Functional Approach To Automated Authorship Analysis”, <i>Journal of Law and Policy</i> , 21(2), 299–315.
CG-7	Coulthard, M. (2000) “Whose Text is it? On the Linguistic Investigation of Authorship”, in S. Sarangi and R.M. Coulthard (eds), <i>Discourse and Social Life</i> , London, Longman, 282–287.
CG-8	Coulthard, M. (2013) “On Admissible Linguistic Evidence”, <i>Journal of Law and Policy</i> , 23(2), 441–466.
CG-9	Grant, T. (2007) “Quantifying evidence in forensic authorship analysis”, <i>International Journal of Speech Language and the Law</i> , 14(1), 1–25.
CG-10	Grant, T. and Baker, K. (2001) “Identifying reliable, valid markers of authorship: a response to Chaski”, <i>Forensic Linguistics</i> , 8(1), 66–79.
CG-11	Grant, T. (2012) “TXT 4N6: Method, Consistency, and Distinctiveness in the Analysis of SMS Text Messages”, <i>Journal of Law and Policy</i> , 21(2), 467–494.
CG-12	Johnson, A. and Wright, D. (2014) “Identifying idiolect in forensic authorship attribution: an n-gram textbite approach”, <i>Language and the Law</i> , 1(1), 37–69.
CG-13	McMenamin, G.R. (2001) “Style markers in authorship studies”, <i>International Journal of Speech Language and the Law</i> , 8(2), 93–97.
CG-14	Nini, A. and Grant, T. (2013), “Bridging the gap between stylistic and cognitive approaches to authorship analysis using Systemic Functional Linguistics and multidimensional analysis”, <i>International Journal of Speech Language and the Law</i> , 20(2), 173–202.
CG-15	Tabachnick, B.G. and Fidell, L.S. (2013) <i>Using Multivariate Statistics</i> , 6 th ed., Boston, Pearson, Chapter 11, ‘Discriminant Function Analysis’, 456–516.
CG-16	Winter, E. and Woolls, D. (1993) ‘Identifying authorship in a co-written novel’, Internal report for the University of Birmingham English Department.